ST2334 Cheatsheet

github.com/reidenong/cheatsheets, AY23/24 S2

1. Probability and Counting

Inverse Probability Formula:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Independent Events

Events A and B are independent ($\perp$) if and only if $P(A \mid B) = P(A)$ or $P(B \mid A) = P(B)$,
ie. $P(A \cap B) = P(A) \cdot P(B)$

Mutually Exclusive

Events A and B are mutually exclusive iff $P(A \cap B) = \emptyset$

2. Random Variables

Probability Mass Function:

For a *discrete* random variable X, the probability mass function (pmf) of X is

$$f(x) : P(X = x), \forall x \in R_X, 0 \text{ otherwise}$$

(1) $f(x_i) \geq 0, \forall x_i \in R_X$
(2) $f(x) = 0, \forall x \notin R_X$
(3) $\sum_{x \in R_X} f(x) = 1$

Probability Density Function:

For a *continuous* random variable X, the probability density function (pdf) of X is a function that satisfies:
(1) $f(x) \geq 0, \forall x \in R_X \wedge f(x) = 0 \forall x \notin R_X$
(2) $\int_{R_X} f(x)dx = 1$
(3) For some $a \leq b$, $P(a \leq X \leq b) = \int_a^b f(x)dx$ (4) $f(x) \geq 0$ but not necessarily ≤ 1.

Cumulative Distribution Function:

For *any* random variable X, the *cdf* of X is defined by

$$F(x) = P(X \leq x)$$

If X is a discrete random variable, then for any two numbers $a < b$, we have

$$P(a \leq X \leq b) = F(X \leq b) - F(X < a) = F(b) - F(a-)$$

where $F(a-) = \lim_{x \uparrow a} F(x)$
= largest value in R_X that is less than a.

Further, $0 \leq F(x) \leq 1$

If X is a continuous random variable, then
$F(x) = \int_{-\infty}^x f(t)dt$,
$P(a \leq X \leq b) = F(b) - F(a)$

CDFs are right continuous, have a maximum value of 1, and non decreasing.

Expectation

For *discrete* X, the expectation of X is $E(X) = \sum_{x \in R_X} x \cdot f(x)$. For *continuous* X, the expectation of X is $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$.
(1) $E(aX + b) = aE(X) + b$
(2) $E(X + Y) = E(X) + E(Y)$
(3) Let g be an arbitrary function.
if X discrete, $E[g(X)] = \sum_{x \in R_X} g(x) \cdot f(x)$.
if X continuous, $E[g(X)] = \int_{R_X} g(x) \cdot f(x)dx$

Variance

$$\sigma_X^2 = V(X) = E\left[(X - \mu_X)^2\right]$$

If X discrete, $V(X) = \sum_{x \in R_X} (x - \mu_X)^2 \cdot f(x)$
If X continuous, $V(X) = \int_{R_X} (x - \mu_X)^2 \cdot f(x)dx$
(1) $\forall X, V(X) \geq 0$. Equality holds when X is constant.
(2) $V(aX + b) = a^2 V(X)$
(3) $V(X) = E(X^2) - [E(X)]^2$
(4) The standard deviation of X is $\sigma_X = \sqrt{V(X)}$

3. Joint Distributions

Discrete Joint Probability Function

$f(x, y) = P(X = x, Y = y), \forall (x, y) \in R_{X,Y}$ Properties:
(1) $f(x, y) \geq 0, \forall (x, y) \in R_{X,Y}$
(2) $f(x, y) = 0$ if $(x, y) \notin R_{X,Y}$
(3) $\sum_{x \in R_X} \sum_{y \in R_Y} f(x, y) = 1$

Continuous Joint Probability Function

$f(x, y) = P(X \leq x, Y \leq y), \forall (x, y) \in R_{X,Y}$ Properties:
(1) $f(x, y) \geq 0, \forall (x, y) \in R_{X,Y}$
(2) $f(x, y) = 0$ if $(x, y) \notin R_{X,Y}$
(3) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$

Conditional Distribution

The **conditional probability function of** Y **given** $X = x$ (the distribution of Y given that $X = x$) is

$$f_{Y|X}(y \mid x) = P(Y = y \mid X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Independent Random Variables

1. Two random variables X and Y are independent iff for all x and y, $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$
2. $R_{X,Y}$ needs to be a product space, ie. $R_{X,Y} = R_X \times R_Y$ for X and Y to be independent.

Checking Independence

(a) $R_{X,Y}$, the range where the probability function is positive, is a product space.
(b) $\forall (x, y) \in R_{X,Y}, f_{X,Y}(x, y) = C \times g_1(x) \cdot g_2(y)$

Expectation and Variance of Random Variables

1. If (X, Y) is a 2-D discrete random variable, then $E[g(X, Y)] = \sum_x \sum_y g(x, y) \cdot f_{X,Y}(x, y)$
2. If (X, Y) is a 2-D continuous random variable, $E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y)dxdy$

Covariance

The covariance of two random variables X and Y is

$$\text{cov}(X, Y) = E[(X - \mu_X) \cdot (Y - \mu_Y)]$$

If discrete,
$\text{cov}(X, Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) \cdot f_{X,Y}(x, y)$
If continuous,
$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f_{X,Y}(x, y)dxdy$

(1) $\text{cov}(X, Y) = E(XY) - \mu_X \mu_Y$
(2) if X, Y are independent, then $\text{cov}(X, Y) = 0$
(3) $\text{cov}(aX + b, cY + d) = ac \cdot \text{cov}(X, Y)$
(4) $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab \cdot \text{cov}(X, Y)$

Variance and Covariance

$V(X + Y) = V(X) + V(Y) + 2 \text{ cov}(X, Y) \Rightarrow$

1. For independent X and Y, $V(X \pm Y) = V(X) + V(Y)$
2. For any random variables $X_1, X_2, ..., X_n$, $V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + 2 \sum_{i<j} \text{cov}(X_i, X_j)$

4. Special Distributions

Uniform Distributions

$$f_X(x) = \frac{1}{k}, x \in \{x_1, x_2, ..., x_k\}$$

1. $\mu_X = \frac{x_1 + x_2 + ... + x_k}{k}$
2. $\sigma_X^2 = E(X^2) - [E(X)]^2 = \left(\frac{1}{k}\right) \sum_{i=1}^k (x_i - \mu_X)^2$

Bernoulli Trial

A Bernoulli trial is a random experiment with two possible outcomes: success (S) and failure (F).
$f_X(x) = p^x (1 - p)^{1-x}$, for $x \in \{0, 1\} = p$
$\Rightarrow \mu_X = p, \sigma_X^2 = p(1 - p)$

Binomial Distribution

In n independent Bernoulli trials,
P(x successes in n trials) = $P(X = x)$ =

$$\binom{n}{x} p^x (1 - p)^{n-x}, \text{for } x = 0, 1, ..., n$$
$$E(X) = np, V(X) = np(1 - p)$$

Negative Binomial Distribution

Let X be the number of independent and identically distributed Bernoulli trials needed until the kth success occurs. Then X follows a negative binomial distribution $X \sim \text{NB}(k, p)$, defined by

$$P(X = x) = \binom{x - 1}{k - 1} p^k (1 - p)^{x-k}, \text{for } x = k, k + 1, ...$$
$$E(X) = \frac{k}{p}, V(X) = \frac{k(1 - p)}{p^2}$$

Geometric Distribution

Let X be the number of independent and identically distributed Bernoulli trials needed until the first success occurs. Then X follows a geometric distribution $X \sim \text{Geom}(p)$, defined by

$$P(X = x) = p(1 - p)^{x-1}$$
$$E(X) = \frac{1}{p}, V(X) = \frac{1 - p}{p^2}$$

Poisson Distribution

The Poisson distribution X denotes the number of events occuring in a fixed region of time or space. $X \sim \text{Poisson}(\lambda)$, where $\lambda > 0$ is the expected number of occurrences in the region.

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \text{for } k = 0, 1, ...$$
$$E(X) = V(X) = \lambda$$

Poisson Process

A Poisson process counts the number of evemts within some interval of time. The defining properties of a Poisson process with rate parameter α are

- Expected number of occurrences in an interval of length t is αt
- There are no simultaneous occurrences
- The number of occurrences in non-overlapping intervals are independent

Then the number of occurrences in any interval T of a Poisson process follows a Poisson distribution
$X \sim \text{Poisson}(\alpha T)$

Poisson Approximation to Binomial

Let $X \sim \text{Bin}(n, p)$. Then as $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains constant, the binomial distribution converges to a Poisson distribution $X \sim \text{Poisson}(np)$.

$$\lim_{n \to \infty; p \to 0} P(X = x) = \frac{e^{-np}(np)^x}{x!}$$

The approximation is good for

- $n \geq 20$ and $p \leq 0.05$
- $n \geq 100$ and $np \leq 10$

Special Continuous Distributions

Uniform Distribution

$X \sim U(a, b)$ follows a uniform distribution over the interval (a, b), if its probability density function is
$f_X(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, 0 otherwise.

$$E(X) = \frac{a + b}{2}, V(X) = \frac{(b - a)^2}{12}$$

The cumulative distribution function is then $F_X(x) = \frac{x-a}{b-a}$, for $a \leq x \leq b$.

Exponential Distribution

Often used to model the waiting time for first success in continuous time, $X \sim \text{Exp}(\lambda), \lambda = \frac{1}{\mu} > 0$ if its probability density function is

$$f_X(x) = \lambda e^{-\lambda x}, \text{for } x \geq 0, 0 \text{ otherwise}$$
$$E(X) = \frac{1}{\lambda}, V(X) = \frac{1}{\lambda^2}$$

The cumulative distribution function of the exponential distribution with parameter λ is then

$$P(X \leq x) = 1 - e^{-\lambda x}, \text{ for } x \geq 0$$
$$P(X > x) = e^{-\lambda x}, \text{ for } x \geq 0$$

In addition, the exponential distribution can be shown to have "no memory", as for any two positive numbers s and t,

$$P(X > s + t \mid X > s) = P(X > t)$$

Normal Distribution

$X \sim N(\mu, \sigma^2)$ if its pdf is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{for } x \in \mathbb{R}$$
$$E(X) = \mu, V(X) = \sigma^2$$

The pdf is symmetric about μ, as σ increases, the curve flattens.

Given $X \sim N(\mu, \sigma^2)$, let $Z = \frac{X - \mu}{\sigma}$. Then $Z \sim N(0, 1)$. In general, if $Z \sim N(0, 1)$, then $\sigma Z + \mu \sim N(\mu, \sigma^2)$.

Z Upper quantile

Let $Z \sim N(0,1)$. Then for any $0 < \alpha < 1$, the α upper quantile of Z is the number z_α such that

$$P(Z > z_\alpha) = \alpha$$

Normal Approximation to Binomial

Let $X \sim \operatorname{Bin}(n,p)$, then as $n \to \infty$,

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

a rule of thumb is to approximate when $np > 5$ and $n(1-p) > 5$, where in general np remains a constant when $n \to \infty \wedge p \to 0$.

We also apply continuity correction.

$P(X = k)$	$P\left(k - \frac{1}{2} < X < k + \frac{1}{2}\right)$
$P(a \le X \le b)$	$P\left(a - \frac{1}{2} < X < b + \frac{1}{2}\right)$
$P(a < X < b)$	$P\left(a + \frac{1}{2} < X < b - \frac{1}{2}\right)$
$P(X \le c)$	$P(0 \le X \le c)$
$P(X > c)$	$P(c < X \le n)$

5. Sampling Distributions

Simple Random Sample

A sample that is chosen such that every subset of n observations of the population has the same probability of being selected.

Statistics

A statistic is a function of the sample data.

Sampling Mean, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sampling Variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

Standard Error (Standard Deviation)

The standard error of $\bar{X}$, $\sigma_{\bar{X}}$,describes how much $\bar{x}$ tends to vary from sample to sample of size n.

Law of Large Numbers

As sample size n increases, $\frac{\sigma_X^2}{n}$ decreases, and $\bar{X}$ approaches μ_X.

$$P\left(\left|\bar{X} - \mu_X\right| > \varepsilon\right) \to 0, \text{as } n \to \infty$$

Central Limit Theorem

If we take the mean of a large number of independent smaples, then the distribution of the mean will be approximately normal.

χ^2 Distribution

Let $Z_1, \ldots, Z_n$ be n independent and identically distributed standard normal random variables. A random variable with the same distribution as $Z_1^2 + \ldots + Z_n^2$ is called a χ^2 random variable with n degrees of freedom, denoted as $\chi^2(n)$.

- $\chi^2(n;\alpha)$ is defined such that for $Y \sim \chi^2(n)$, $P(Y > \chi^2(n;\alpha)) = \alpha$.

1. if $Y \sim \chi^2(n)$, $E(Y) = n$, $V(Y) = 2n$
2. if $Y_1 \sim \chi^2(n_1)$ and $Y_2 \sim \chi^2(n_2)$, then $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$
3. For large n, $\chi^2(n)$ is approximately $N(n, 2n)$.

Theorem 12

If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the random variable

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2} \sim \chi^2(n-1)$$

t-Distribution

Suppose $Z \sim N(0,1)$ and $U \sim \chi^2(n)$. If Z and U are independent, then $T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n$.

$$\text{for } T \sim t_n, P(T > t(n;\alpha)) = \alpha$$

1. When $n \ge 30$, we can replace it by $N(0,1)$
2. If $T \sim t(n)$, $E(T) = 0$ and $V(T) = \frac{n}{n-2}$

F-Distribution

Suppose $U_1 \sim \chi^2(n_1)$ and $U_2 \sim \chi^2(n_2)$. If U_1 and U_2 are independent, then $F = \frac{\left(\frac{U_1}{n_1}\right)}{\left(\frac{U_2}{n_2}\right)} \sim F(n_1, n_2)$.

$F(n,m;\alpha) = k \Rightarrow P(F > k) = \alpha$, where $F \sim F(n,m)$

1. $E(X) = \frac{m}{m-2}$, $V(X) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$
2. $F \sim F(n,m) \Rightarrow \frac{1}{F} \sim F(m,n)$
3. $F(n,m;\alpha) = \frac{1}{F(m,n;1-\alpha)}$

6. Estimation

An estimator is a rule, usually expreseed as a formula, that tells us how to calculate an estimate based on the data in a sample.

Maximum error of estimate

$E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

7. Hypothesis Testing

Errors

	Do not reject H_0	Reject H_0
H_0 true	Correct	**Type I Error**
H_0 false	**Type II Error**	Correct

Level of Significance

LoS (α) is the probability of making a Type I error.

$$\alpha = P(\text{Type I Error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

Power of a Test

Power of a test is the probability of correctly rejecting a false null hypothesis.

$$\beta = P(\text{Type II Error}) = P(\text{Do not reject } H_0 \mid H_0 \text{ is false})$$

$$\text{power} = 1 - \beta = P(\text{Reject } H_0 \mid H_0 \text{ is false})$$

p-value

The p-value is the probability of obtaining a test statistic at least as extreme than the observed sample value given H_0 is true, if $p < \alpha$, reject H_0.

Test statistics for Paired samples

- Each X_i is dependent on each Y_i, but each pair is independent of all other pairs.
- Define $D_i = X_i - Y_i$. Then D is a random sample from a population with mean μ_D and variance σ_D^2.

Test statistics for Population Mean

Case	Population	σ	n	Confidence Interval	Test Statistic
I	Normal	known	any	$\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
II	any	known	≥ 30	$\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
III	Normal	unknown	< 30	$\bar{x} \pm t_{n-1;\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$
IV	any	unknown	≥ 30	$\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim N(0,1)$

Test statistics for Independent samples

Population	Variance	σ_1, σ_2	n	Confidence Interval	Test Statistic
any	known	unequal	≥ 30	$(\bar{x} - \bar{y}) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
Normal	known	unequal	any	$(\bar{x} - \bar{y}) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
any	unknown	unequal	≥ 30	$(\bar{x} - \bar{y}) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$
Normal	unknown	equal	< 30	$(\bar{x} - \bar{y}) \pm t_{n_1+n_2-2,\frac{\alpha}{2}} \cdot s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\frac{(\bar{X} - \bar{Y})}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$
any	unknown	equal	≥ 30	$(\bar{x} - \bar{y}) \pm Z_{\frac{\alpha}{2}} \cdot s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\frac{(\bar{X} - \bar{Y})}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$

* Variance is assumed to be equal if $\frac{1}{2} < \frac{s_1}{s_2} < 2$.

Pooled Estimator

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$